Lyon, M., Munafo, M., Gaunt, T., & Davey Smith, G. (2019). Availability of public databases for triangulation of findings. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(32), 15766-15767. https://doi.org/10.1073/pnas.1908893116

LETTER

Availability of public databases for triangulation of findings

Matthew Lyon[a,b,1], Marcus Munafo[a,b], Tom Gaunt[a,b], and George Davey Smith[a,b]

Struk et al. (1) report evidence that the rs13499 single-nucleotide polymorphism (SNP) in the *PRKG1* gene is associated with foraging and goal pursuit behavior. The authors used an expression quantitative trait locus (eQTL) for *PRKG1* obtained from the CommonMind Consortium ($n = 467$) to test for association between gene expression and phenotype (2). However, the original SNP–gene expression association had an opposite direction of effect (increasing allele A in open-access CommonMind analysis vs. increasing allele C reported by Struk et al.; Table 1). Furthermore, the eQTL was not robustly associated after taking account of multiple testing (false detection rate [FDR] between 0.05 and 0.1) and did not replicate in BrainSeq (ref. 3; $n = 412$), casting doubt on both the robustness and the interpretation of the reported results (Table 1).

Importantly, the sample size (sample 1 $n = 153$; sample 2 $n = 437$) was too small to reliably detect effect sizes anticipated with polygenic inheritance (4). Low power increases type II errors (false negative) and the likelihood that a "statistically significant" finding represents a type I error (false positive). The latter is often exacerbated by publication bias. When true effects are present their size estimates are often inflated (5). Struk et al. (1) report very weak statistical evidence for association of rs13499 and regulatory mode predominance (sample 1 $P = 6.7 \times 10^{-2}$; sample 2 $P = 7.0 \times 10^{-3}$), which would have been considered inadequate even in the heyday of candidate gene studies (6).

One approach to improve rigor with limited sample size is through triangulation—for example by comparing the results of smaller studies using narrow (e.g., mechanistic) phenotypes with those from larger studies using broader phenotypes. This latter step can now be easily done using publicly available databases such as the MR-Base database and accompanying phenome-wide association study browser (refs. 7 and 8; http://phewas.mrbase.org). This resource contains full summary genetic association statistics for 22,311 traits with many analyses on more than 400,000 participants.

Stringent filtering of CommonMind eQTLs (dataset syn5652289) identified a single SNP (rs1904701) influencing *PRKG1* expression (FDR < 0.01; pruning threshold $R^2 < 0.001$) with consistent directionality in BrainSeq ($P = 1.18 \times 10^{-15}$). We tested association of both SNPs (rs1904701 and rs13499, correlation $R^2 = 1.8 \times 10^{-3}$) against a series of traits one would expect to correlate with goal pursuit behavior (Table 2). We found no robust evidence of association with *PRKG1* expression and foraging behavior and therefore no clear support for the findings of Struk et al. (1).

Candidate gene studies have a poor track record with respect to producing robust, replicable findings (9), particularly in contrast with genome-wide association studies. At the very least, candidate gene studies should now report associations with broad phenotypes measured in large population samples, given the ease with which such results can be obtained from publicly available datasets.

Acknowledgments

This study was funded by the NIHR Biomedical Research Centre at University Hospitals Bristol National Health Service Foundation Trust and the University of Bristol. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. M.L., M.M., T.G., and G.D.S. work in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol, which is supported by the Medical Research Council and the University of Bristol (MC_UU_00011/1, MC_UU_00011/4, MC_UU_00011/7).

[a]National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University of Bristol, Oakfield House, Bristol BS8 2BN, United Kingdom; and [b]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Oakfield House, Bristol BS8 2BN, United Kingdom

Author contributions: M.M. and G.D.S. designed research; M.L. performed research; M.L., M.M., T.G., and G.D.S. analyzed data; and M.L., M.M., T.G., and G.D.S. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: matt.lyon@bristol.ac.uk.

Published online July 23, 2019.

Table 1. Association between rs13499 and *PRKG1* expression in dorsolateral prefrontal cortex tissue

Study	Increasing allele	Decreasing allele	No. of samples	*P*
Struk et al. (1)	C	A	600	2.32×10^{-3}
CommonMind adjust −SVA (syn5652281)	NA	NA	467	FDR > 0.2
CommonMind adjust +SVA (syn5652289)	A	C	467	FDR between 0.05 and 0.1
BrainSeq	NA	NA	412	NS

Increasing and decreasing allele refer to *PRGK1* expression levels. NS, nonsignificant. FDR, false detection rate. NA, not applicable. +/− SVA, with or without adjustment for surrogate variable analysis. CommonMind open-access eQTL analysis (obtained 10/05/19) exact *P* values are not provided. BrainSeq eQTL analysis (obtained 10/05/2019) exact *P* values are not provided for nonsignificant associations. BrainSeq association significance criteria: FDR < 0.01 with consistent gene expression effect directionality in CommonMind and GTEx and association *P* < 0.01 in CommonMind.

Table 2. Goal pursuit behavior trait association with *PRKG1* expression quantitative trait locus in MR-Base

Trait	No. of samples	*P*	Cases	Controls	SNP
Distance between home and job workplace	221,909	0.72	NA	NA	rs13499
Distance between home and job workplace	221,909	0.86	NA	NA	rs1904701
Financial situation satisfaction	152,153	0.43	NA	NA	rs13499
Financial situation satisfaction	152,153	0.86	NA	NA	rs1904701
Loneliness, isolation	455,364	0.46	82,436	372,928	rs13499
Loneliness, isolation	455,364	0.04	82,436	372,928	rs1904701
Neuroticism score	374,323	0.22	NA	NA	rs13499
Neuroticism score	374,323	0.10	NA	NA	rs1904701
Risk taking	446,279	0.76	117,515	328,764	rs13499
Risk taking	446,279	0.03	117,515	328,764	rs1904701
Work/job satisfaction	105,358	0.01	NA	NA	rs13499
Work/job satisfaction	105,358	0.17	NA	NA	rs1904701

rs13499, SNP identified by Struk et al. (1), rs1904701, independent analysis of CommonMind eQTL dataset. NA, not applicable. Full data available from http://phewas.mrbase.org/snp/rs13499/all/.

1 A. A. Struk *et al.*, Self-regulation and the *foraging* gene (*PRKG1*) in humans. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4434–4439 (2019).
2 M. Fromer *et al.*, Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
3 BrainSeq: A Human Brain Genomics Consortium, BrainSeq: Neurogenomics to drive novel target discovery for neuropsychiatric disorders. *Neuron* **88**, 1078–1083 (2015).
4 D. Altshuler, M. J. Daly, E. S. Lander, Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
5 K. S. Button *et al.*, Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
6 H. M. Colhoun, P. M. McKeigue, G. Davey Smith, Problems of reporting genetic associations with complex outcomes. *Lancet* **361**, 865–872 (2003).
7 G. Hemani *et al.*, The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
8 B. L. Elsworth *et al.*, MRC IEU UK Biobank GWAS pipeline version 1. https://data.bris.ac.uk/data/dataset/2fahpksont1zi26xosyamqo8rr. Accessed 10 May 2019.
9 R. Border *et al.*, No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *Am. J. Psychiatry* **176**, 376–387 (2019).